



This work is supported by the Spanish Project TIN2010-17550 for the **BabelData** project, and by **BNE**



Status Quo and (current) limitations of Library Linked Data

Asunción Gómez-Pérez¹, Philipp Cimiano² and
Daniel Vila-Suero¹

¹Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo sn, 28660 Boadilla del Monte, Madrid

<http://www.oeg-upm.net>

asun,dvila@fi.upm.es

²AG Semantic Computing, CITEC, University of Bielefeld
cimianog@cit-ec.uni-bielefeld.de

Acknowledgments: BNE team, DNB team, Uldis Bojars (NLL), Jodi Schneider (NUIG) and others

SWIB12 – Cologne, 28-11-2012

- The Library Linked Data Cloud
- A (library) Linked Data Life-cycle
- A collection of current limitations
- Conclusions



Library Linked Data cloud

Library Linked Data Cloud

LIBRIS

HJÄLP | IN ENGLISH | ANPASSA | MINA BIBLIOTEK | RENSA HISTORIK



Start

Utökad sökning

Bläddra ämnesvis

Index A-Ö

Boolesk

Deldatabaser

Sökhistorik



ASK A LIBRARIAN

DIGITAL COLLECTIONS

LIBRARY CATALOGS

Search

Search Loc.gov

GO

The Library of Congress > Linked Data Service

LIBRARY OF CONGRESS
LINKED DATA SERVICE

LC Linked Data Service
Authorities and Vocabularies



- Linked D
- About
- Dataset
- Search
- Download
- Technical
- Contact

VIAF

Fichero de Autoridades Virtual Internacional

Búsqueda

Seleccione campo

Todos los nombres

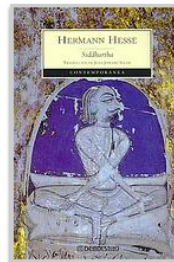
Seleccione índice

Todo VIAF

Términos de búsqueda

hermann hesse 1877 07 02-1962 08 09

Búsqueda



Hesse, Hermann, 1877-1962

Hesse, Hermann

hermann hesse 1877 07 02-1962 08 09

Гессе, Герман, 1877-1962

هسه, هرمان, 1962-1877

הסה, הרמן, 1877-1962

Hesse, Hermann German author and painter, 1877-1962

VIAF ID: 41841418 (Personal)

Permalink: <http://viaf.org/viaf/41841418>

Formas preferidas

100 0 _1a hermann hesse _td 1877 07 02-1962 08 09

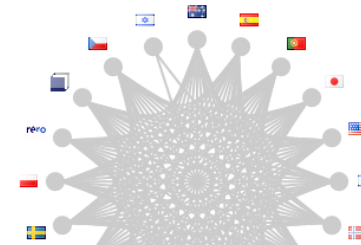
100 1 _1a هسه, هرمان, _td 1877-1962

100 1 _1a Hesse, Hermann, _td 1877-1962

100 1 _1a Hesse, Hermann

100 1 _1a Hesse, Hermann, _td 1877-1962

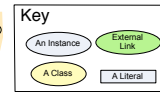
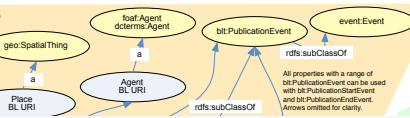
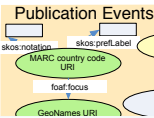
100 1 _1a Hesse, Hermann, _td 1877-1962



Library Linked Data Cloud

British Library Data Model - Book

@prefix bit: <http://www.bl.uk/schemas/bibliographic/blterms#> .
 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
 @prefix owl: <http://www.w3.org/2002/07/owl#> .
 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
 @prefix dct: <http://purl.org/dc/terms/> .
 @prefix skos: <http://www.w3.org/2004/02/skos/core#> .
 @prefix bibo: <http://purl.org/ontology/bibo/> .
 @prefix rda: <http://rda.info/elements/rda/> .
 @prefix geo: <http://purl.org/ontology/geo/> .
 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
 @prefix event: <http://purl.org/event/> .
 @prefix geo: <http://www.w3.org/> .



Author

(BnF) Bibliothèque nationale de France

data.bnf.fr

rechercher dans data.bnf.fr

À propos | Contact

DEUTSCHE
NATIONAL
BIBLIOTHEK

English | Kontakt | A-Z | Förderer | Datenschutz | Impressum | Hilfe | Mein Konto

LEIPZIG
FRANKFURT AM MAIN

RDF/XML | Turtle | Simple JSON | Simple XML | html



Datos enlazados BNE

Personas | Obras | Expresiones | Títulos

Resultados de búsqueda

Mostrar campos de búsqueda

En esta página

?

XX1190180

citation note

(Richa

Seattle

(Richa

Richa

enlaces externos

http://

13660

Richa

14806

source consulted

http://

Interne

tiene fechas de persona

1930-2

tiene nombre de persona

Heffron, Richard T.

tipo

Person / Persona

New version of datos.bne.es by beginning of 2013 including:

- Links to digital objects
- More links to external datasets
- APIs and improved documentation

> XX1190180
 > XX947766
 > XX948211

source consulted
 tiene calificativo de persona
 tiene fechas de persona
 tiene nombre de persona
 tipo

Library Linked Data Cloud

The image shows two overlapping web pages. The top page is the Europeana Professional website, featuring a navigation bar with links like Home, About us, News & Events, Projects, Europeana Foundation, Europeana Network, Provide data, and Re-use data. The bottom page is the CultureSampo portal, titled 'Finnish Culture on the Semantic Web 2.0 (the portal is only partially available in English, sorry)'. It includes a search bar, a sidebar with links like 'Re-use data', 'API services', and 'Linked Open', and a main content area with sections for 'Map Browse and Search', 'Relation Search', 'Search and Organize', 'Collections', and 'Finnish History 1860-1919'. A red box highlights the text 'And many others (see http://thedatahub.org/group/lld)'.

And many others (see <http://thedatahub.org/group/lld>)

Esperando s7.addthis.com...

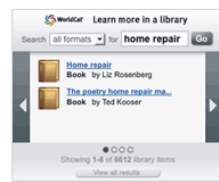
Last Viewed:

video : Kätevät: hiihango



Usar herramientas de WorldCat

Tener los recursos bibliotecarios próximos desde su explorador o página web personalizada. Agregar las bibliotecas del mundo a su móvil y aplicaciones FB.



Widgets y servicios web

[WorldCat Search API](#)

[Aplicaciones y plug-ins del navegador](#)

[Enlazar a WorldCat](#)

[Agregar una chapa WorldCat](#)

Construir una bibliografía

Conseguir citas de materiales bibliotecarios en cinco estilos comunes, y exportarlas a una variedad de formatos tal como EndNote, Reference Manager y RefWorks.



Videotutorial

[Cómo usar citas de WorldCat \(YouTube\)](#)

Duración:1:47

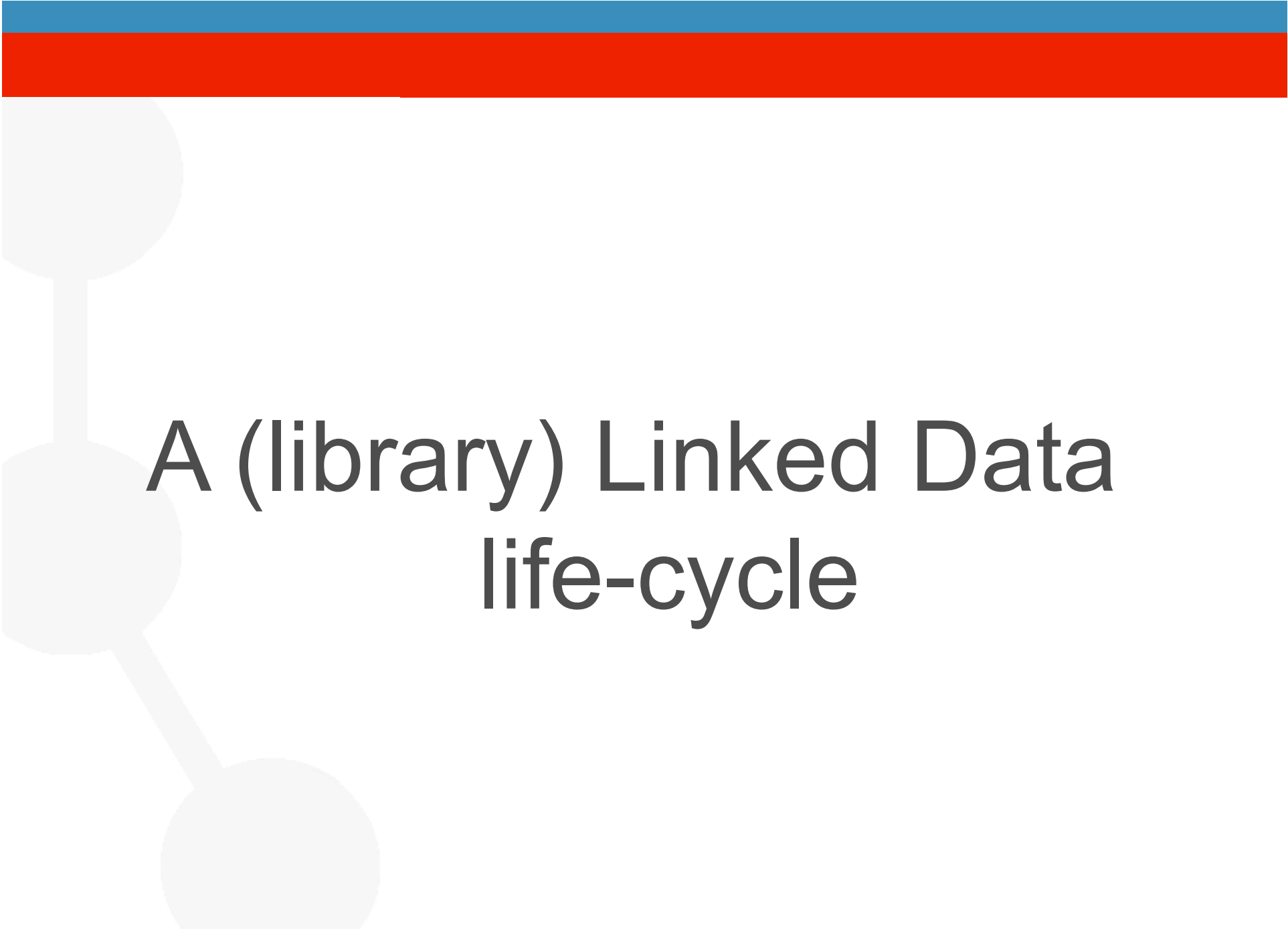


Bibliotecas más populares



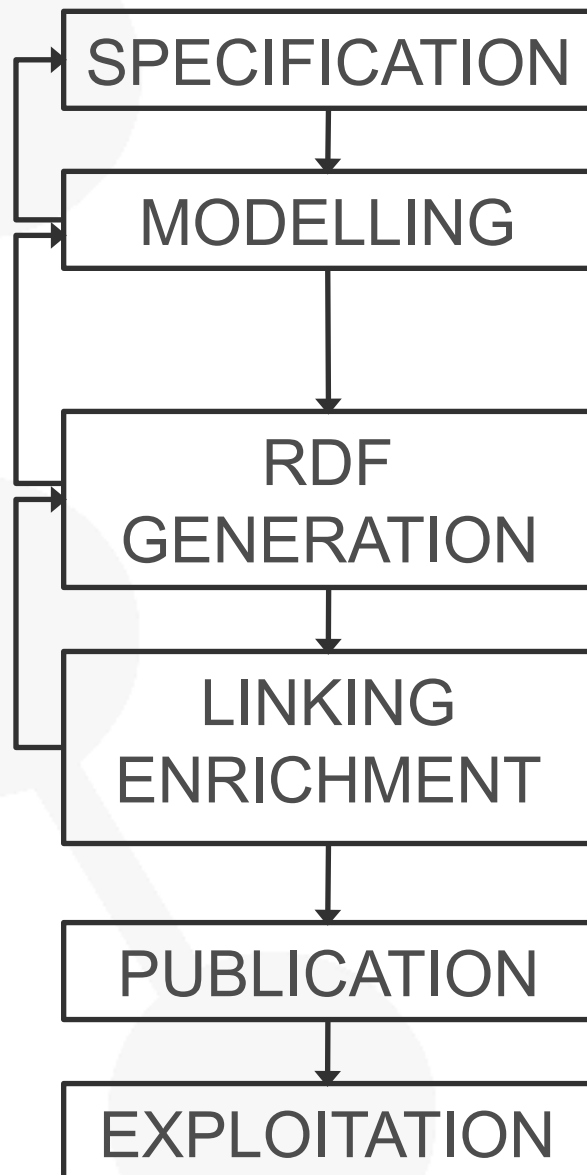
Etiquetas más populares

- **Availability** of Library Linked Data is already a reality
- Many **serious efforts**: id.loc.gov, VIAF, DNB, data.bnf.fr, Bibliographic Framework Initiative etc.
- Still **several challenges** and **limitations** preventing LLD full potential



A (library) Linked Data life-cycle

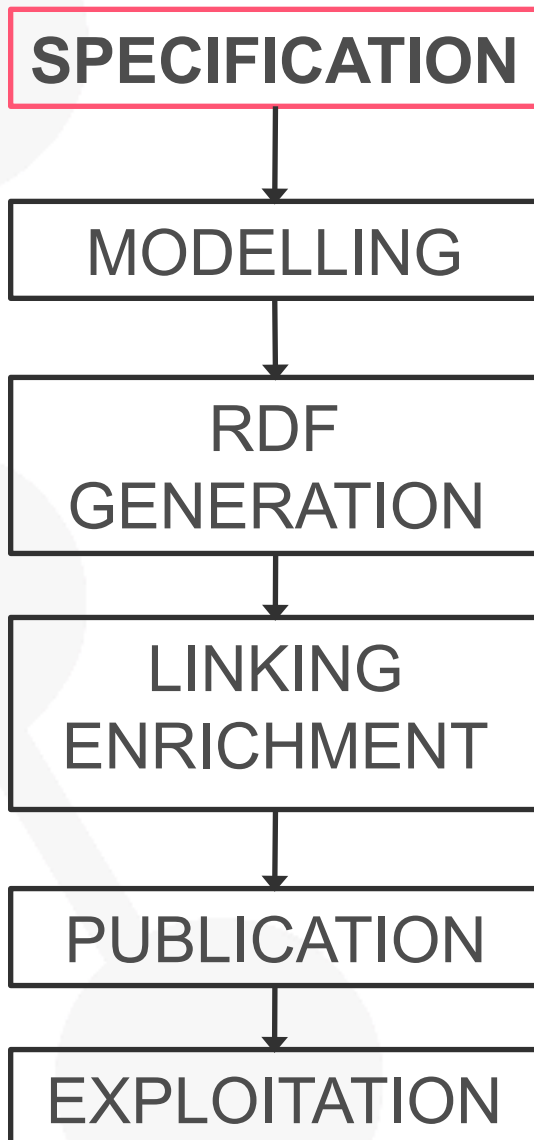
A (Library) Linked Data Life-cycle



- A series of steps or phases
- Based on *Villazón-Terrazas et al.**
- **Others:** LOD2, Datalift, etc. (see http://www.w3.org/2011/gld/wiki/GLD_Life_cycle)

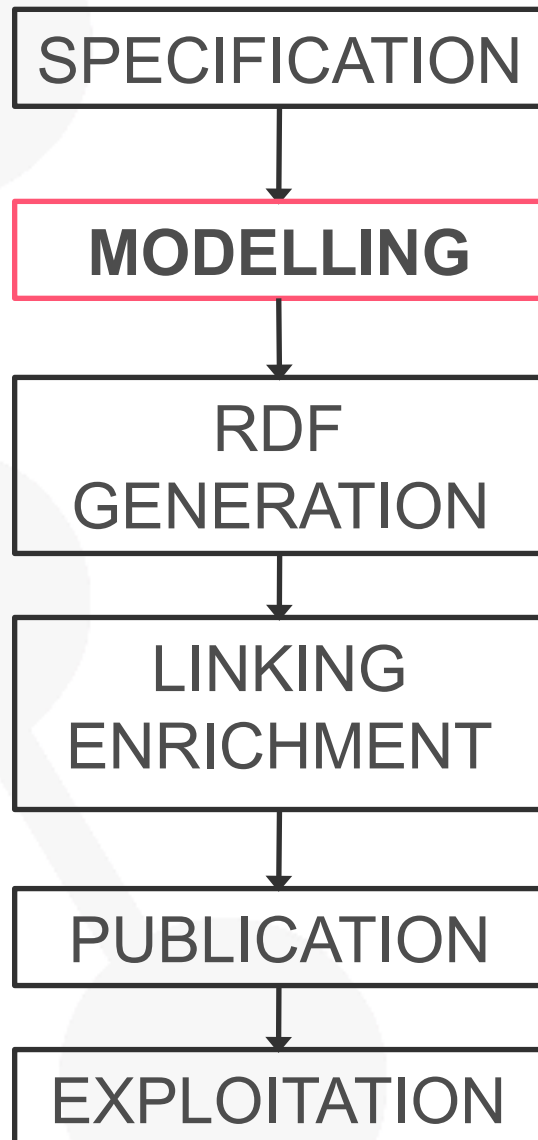
* *Methodological Guidelines for Publishing Government Linked Data*, Boris Villazón-Terrazas, Luis. M. Vilches-Blázquez, Oscar Corcho and Asunción Gómez-Pérez
Linking Government Data Book

A (Library) Linked Data Life-cycle



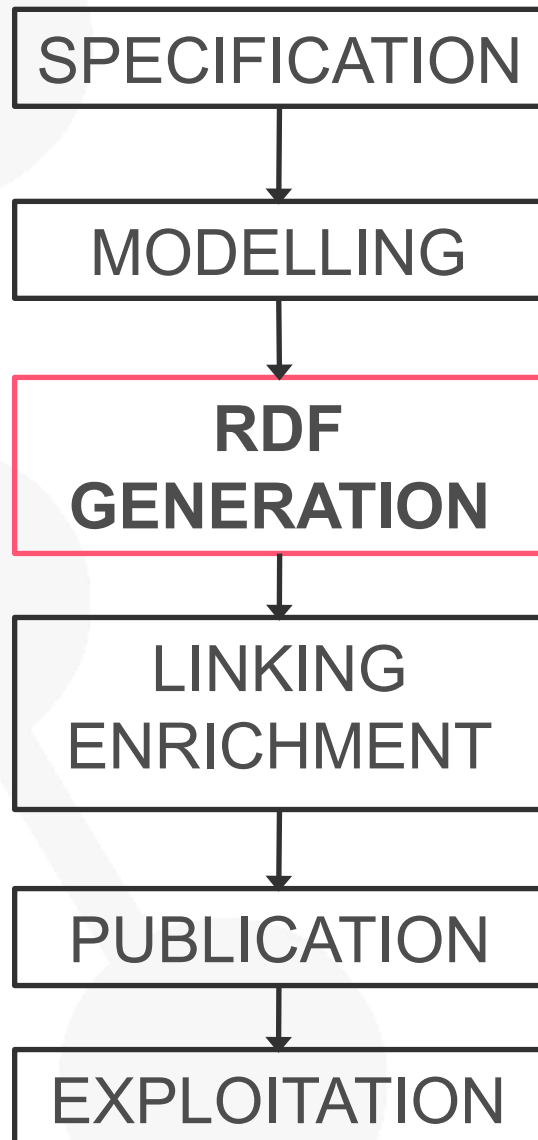
Definition and analysis of source data and their format, structure, etc.

A (Library) Linked Data Life-cycle



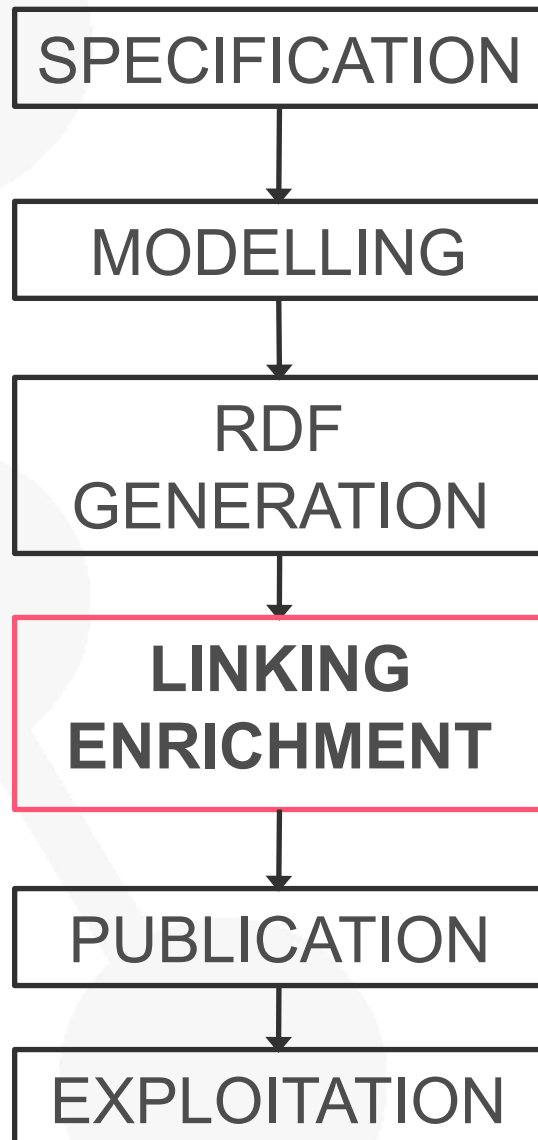
- Selection and reuse of vocabularies to represent LD.
- Creation of new local terms and mapping to existing vocabularies

A (Library) Linked Data Life-cycle



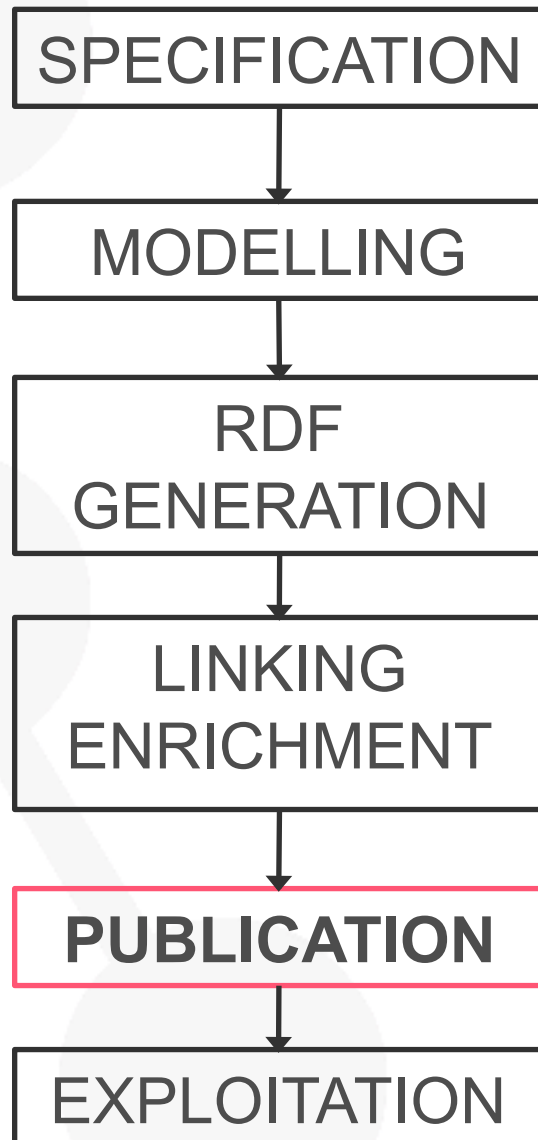
Taking source data, and vocabularies:
Mapping (cross-walk) source data to produce RDF

A (Library) Linked Data Life-cycle



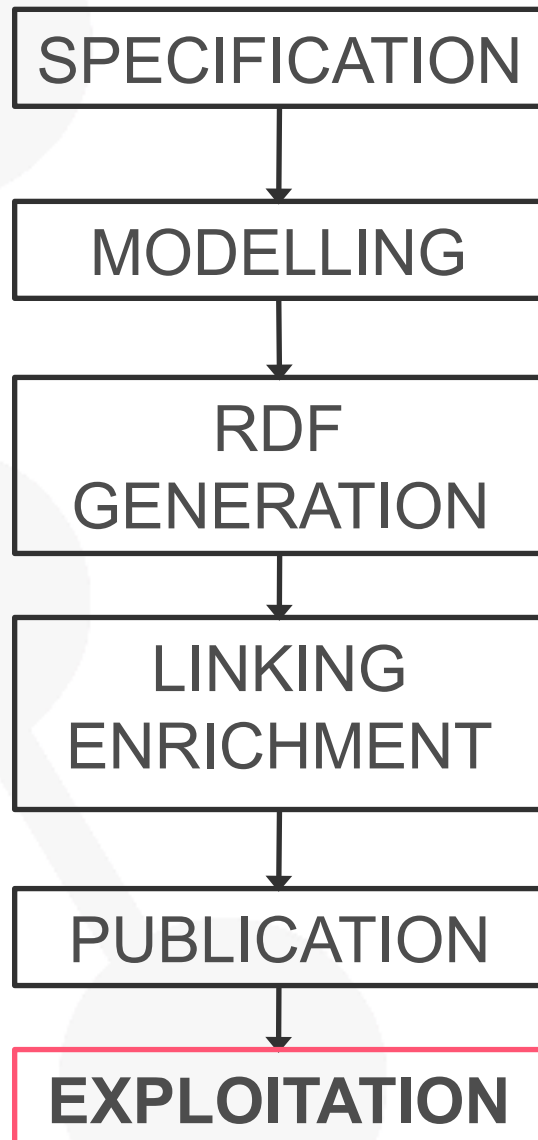
- Discover related resources (ideally in RDF form).
- Enrich RDF data with data from other sources (e.g. substitute literals by URIs in other datasets)

A (Library) Linked Data Life-cycle



- Setup the infrastructure to expose your data to the Web (SPARQL, APIs, dumps).
- Enable discovery of your data (sitemap, voID)
- Include data provenance, license, etc.

A (Library) Linked Data Life-cycle



- Produce user interfaces that integrate your data (and other sources)
- Provide innovative services on top of the data
- Integrate with existing services (e.g. patron services), etc.



A collection of current limitations

SPECIFICATION

- MANY source formats, encodings and schemas:

MODELLING

RDF
GENERATION

LINKING
ENRICHMENT

PUBLICATION

EXPLOITATION

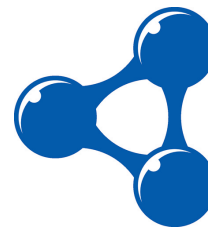


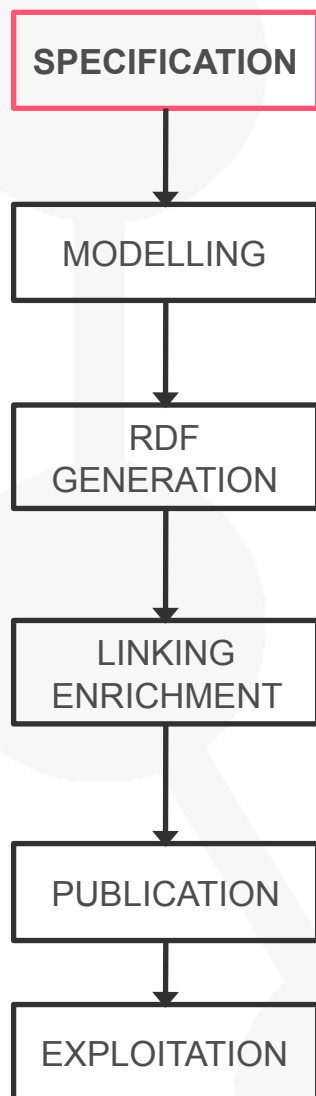
PICA+

MAB

<XML>

CSV





LIMITATIONS:

1. Lack of principled *methods, techniques and tools* to deal with **heterogeneous source formats, schemas and encodings**



metastream, metamorph APIs from culturegraph

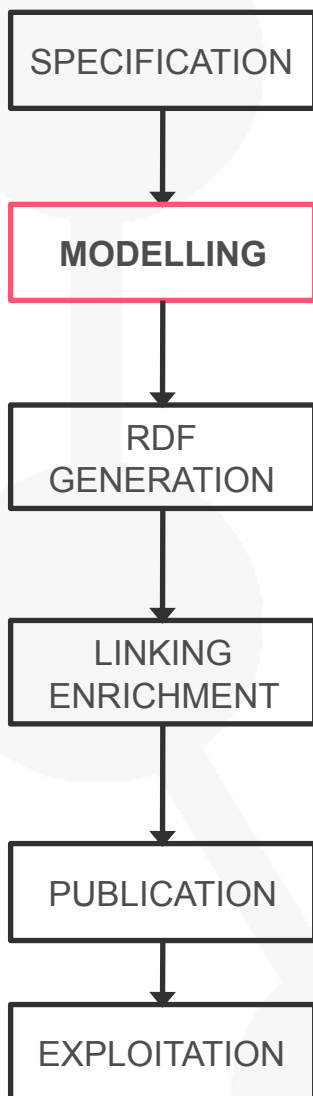
2. Need for **analysis of the semantics of metadata schemas** and the **variation of their usage** accross libraries:



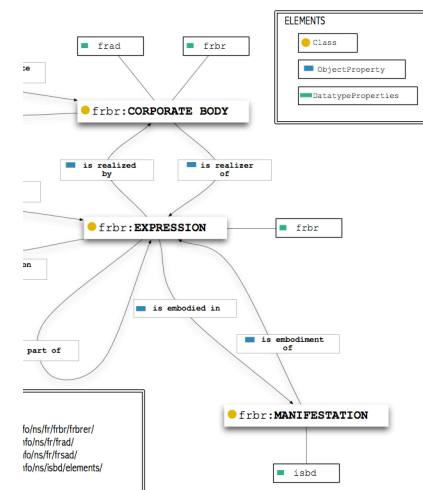
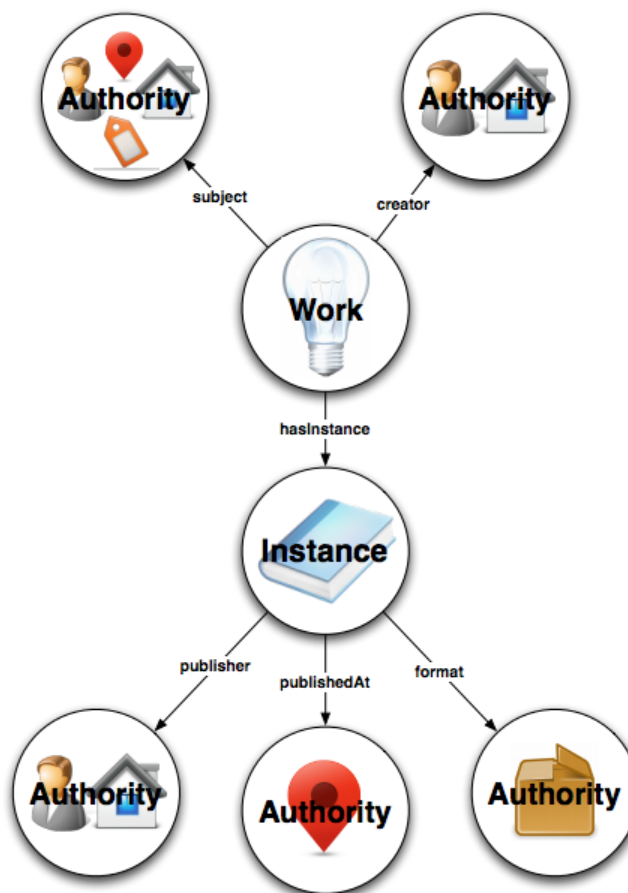
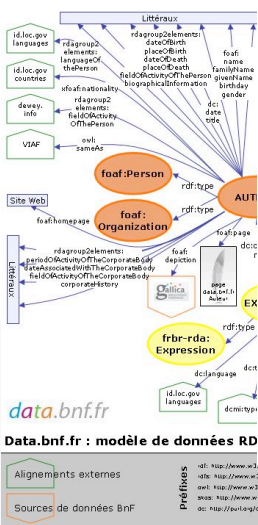
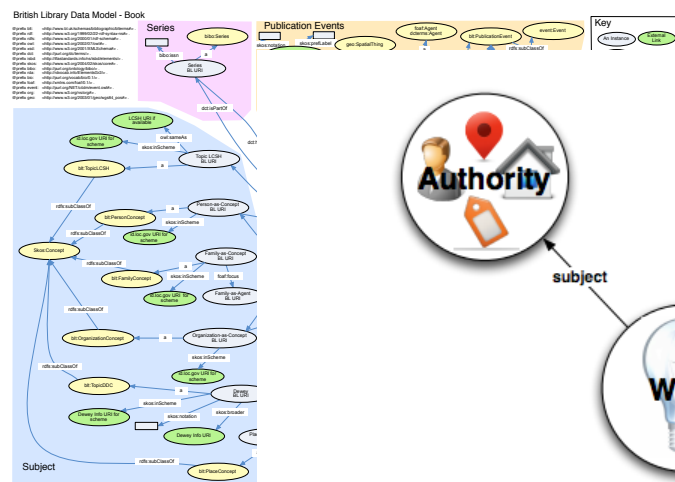
MARC 21: "*Marc21 as Data: A start*".

Karen Coyle's code4lib's article*

* <http://journal.code4lib.org/articles/5468>



• Many different models and approaches



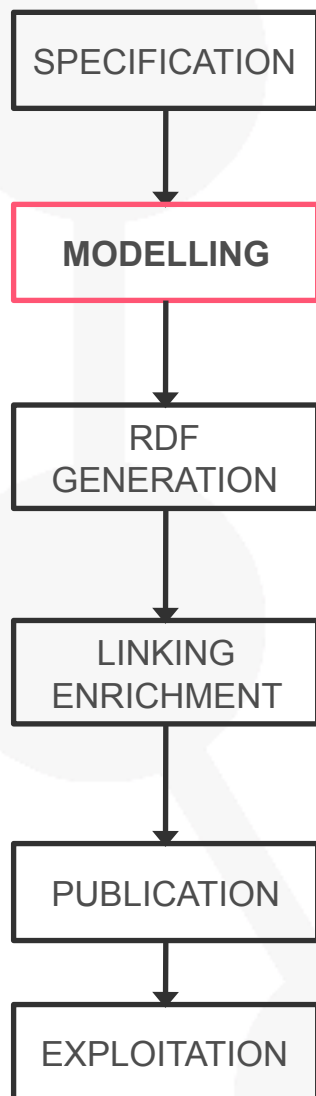
DEUTSCHE
NATIONALE
BIBLIOTHEK

b.de, subscribe to the list and visit archive here)

.license. This copyright applies to the GND Ontology documentation and the ID uses W3C's RDF technology, an open Web standard that can be freely used

of named properties and classes using W3C's RDF technology.

GND stands for "Gemeinsame Normdatei" (Integrated Authority File) and offers a broad range of elements to describe authorities. The GND



LIMITATIONS:

1. Difficulties in **using vocabularies that adapt to**
 - Past and current cataloguing practices
 - Past and current library formats and schemas

2. **Mapping and managing vocabularies**

1. Lack of mapping accross vocabularies
2. High manual effort and costly process



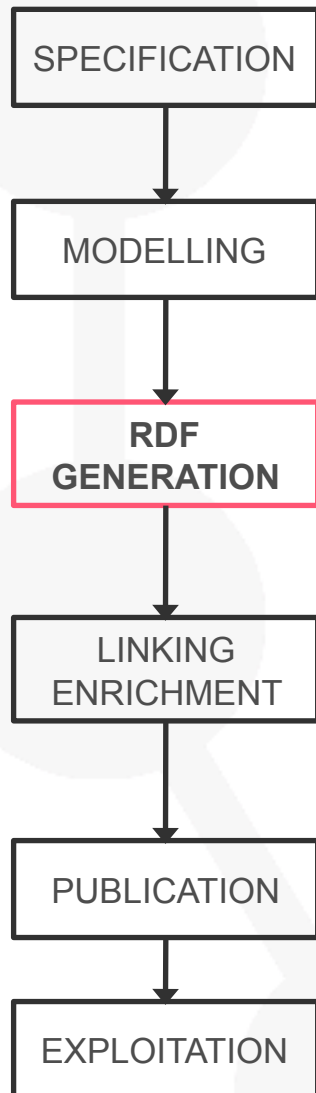
GND ontology,

Dunsire et al "Linked Data vocabulary management"*

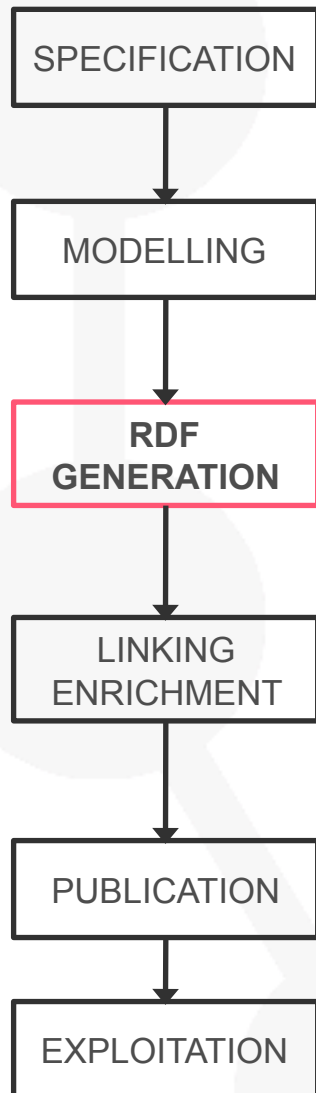
3. Lack of **multilingual** vocabulary elements **description**



IFLA Namespace Multilingual efforts



- Several tools, APIs and systems
- Almost each new project follows its own approach → it is still a costly process
- Participation of library experts is crucial (those who know the formats, e.g. MARC 21)
- There is not a generic mapping from source metadata schemas (due to the differences in use accross libraries, different cataloguing practices)

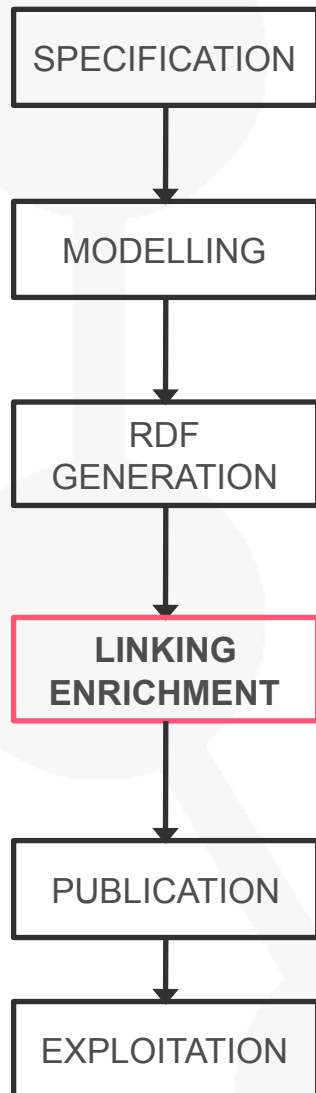


LIMITATIONS:

1. Lack of tools and services that are **easy to use by non-technical users** (experts in library formats)
2. Lack of **integrating tools and services** that provide a **full view** of the mapping and RDF generation process
 - **Analytics of the source data** (e.g. how often is an specific MARC subfield used, how many records will an specific transformation rule affect..)
 - **Dashboard-like** generation that helps to understand how the RDF data has been produced and allows to refine the mappings, find errors, etc.

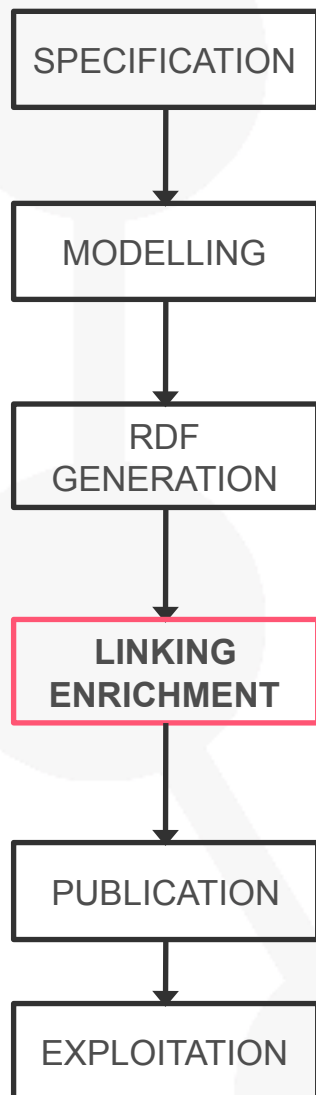


<http://marimba4lib.com>



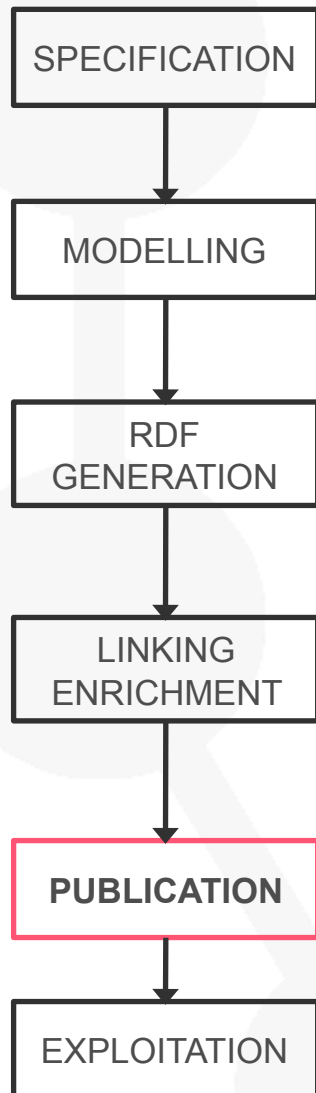
- Already a **high number of links at the authority level** (mostly Persons and Corporate Bodies)
 - VIAF
 - Culturegraph Authorities
 - DNB, BNF, BNE, BL
- Very positive efforts:
 - National library cataloguers adding links during the cataloguing process (e.g. DNB, BNE)
 - Cross-library collaboration for linking (British Library and DNB*)

* <http://www.culturegraph.org/Subsites/culturegraph/DE/Home/news4.html>



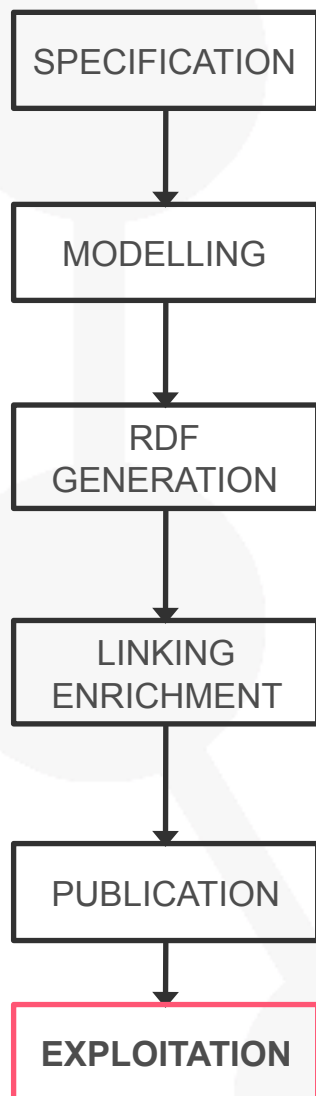
LIMITATIONS:

1. Very **limited linking at the bibliographic level**
2. Lack of **cross-lingual mechanisms** to link resources accross libraries
3. Lack of the **a solid infrastructure to enable links sharing and exchange** accross libraries
4. No **semantic enrichment of the content** (textual, sound and visual) and **linkage** of this content **to the URIs** that represent the real-world entities



LIMITATIONS:

1. No extensive use of mechanisms to indicate provenance, license, last-update, etc. in a per-resource basis
2. Low usage of mechanisms to enable efficient discovery and usage like **void descriptions**
3. Need for scalable and generic infrastructure to facilitate consumption of linked data:
 1. Most of the APIs are not extensively documented
 2. Not every dataset provides search over the data
 3. W3C Linked Data platform WG proposition is promising (REST access to resources and containers: paging, etc.)



LIMITATIONS:

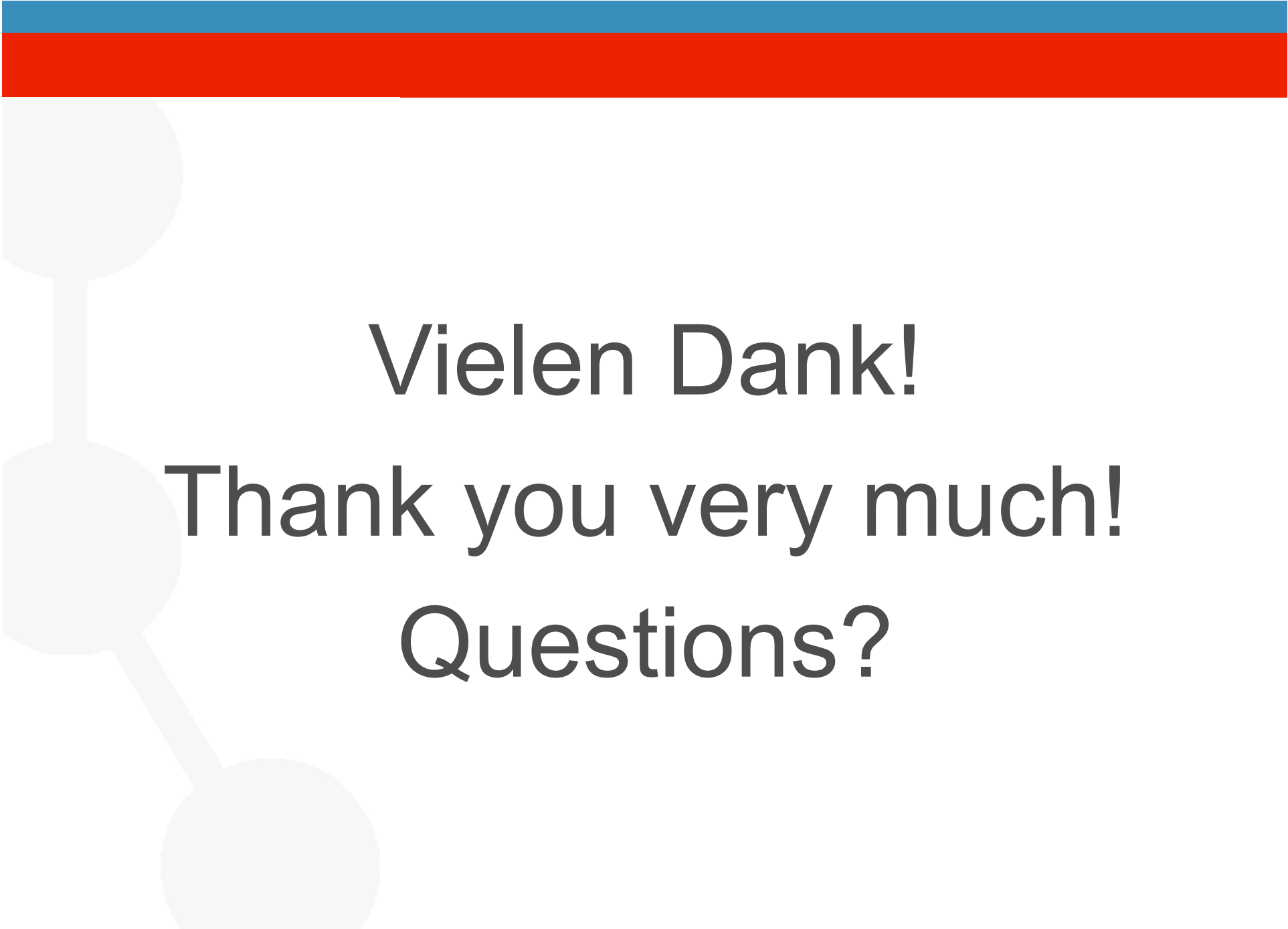
1. Lack of integration of library linked data in
 - Library curation and cataloguing workflows
 - Existing library systems (e.g. digital library systems, patron services)
2. Need of end-user interfaces providing enriched information spaces
 - that integrate several LD sources,
 - allowing for serendipitous discovery and enhanced navigation



CultureSampo,
<http://uilld2013.linkeddata.es> Call for participation!

Conclusions

- There still exist **some barriers**:
 - **Organizational**: Infrastructure costs, integration of linked data into library processes, cross-library collaboration
 - **Language**: Multilingual services, cross-lingual linking and data integration
 - **Data formats and modalities**: Different digital objects and representations are rarely interlinked, cope with heterogeneity of formats and vocabularies
- **But we are on the right path!**



Vielen Dank!
Thank you very much!
Questions?